The Power MOSFET, A Breakthrough in Power Device Technology

Recent advances in the technology of fabricating MOSFET's have enabled manufacturers to break free of the power limitations of the early devices. While the power MOSFET is a relative newcomer to the power switch field, it promises exciting performance advantages over the more conventional bipolar transistor. By providing much higher switching speeds (a few nanoseconds is typical), high input impedance with low drive requirements, simplified multi-device operation, and greatly improved safe operating area for power switching applications, the power MOSFET will replace the bipolar transistor in many applications, and provide new circuit opportunities that do not exist with bipolar technologies. Several different types of power MOSFET's are already available, and the immediate future promises many more devices with steadily increasing performance.

This applications note is intended to provide the design engineer with a basic understanding of how power MOSFET's work, the structures presently in use, the electrical characteristics of the devices and some guidance on application.

STRUCTURES

At the present time, there are a host of different and often confusing names for power FET devices, such as VMOS, VFET, HEXFET®, TMOS, DMOS, ZMOS, etc. In actuality there are only three basic structures: vertical junction FET's, V-groove MOSFET's and vertical DMOS FET's.

Planar MOSFET

Figure 1 is a cross section of a conventional planar N-channel enhancement mode MOSFET. Fabrication of this device begins with a P substrate into which N^+ regions are diffused. SiO_2 is then grown and etched for the aluminum which is deposited to form the source, gate and drain connections. If no bias is applied to the gate, this device is essentially two back-to-back PN diodes and no conduction occurs. If, however, the gate is made positive with respect to the source, the electrostatic field draws electrons near the surface of the P region. This inverts that region to N material, and a channel is formed, allowing conduction between source and drain. Note that the MOSFET is a majority carrier device which acts like a voltage controlled resistor during conduction. The result is an extremely fast switch with no storage time effects. There are, however, a number of drawbacks for high power use which, for practical purposes, eliminate this structure from consideration for high power use:

1. The length of the channel is controlled by the mask spacing of the N^+ regions. Due to the limits of accuracy of photomask technology, it is necessary to have relatively wide spacing. This produces long channel lengths which increase the ON resistance for a given area of silicon.
2. The source, gate and drain conductors are all on the same surface. This metallization takes up a major portion of the die area, further increasing the ON resistance.
3. This structure has large inherent capacitances, for its current capability, especially gate to drain. This reduces the gain bandwidth and increases the drive power in repetitive pulse applications.

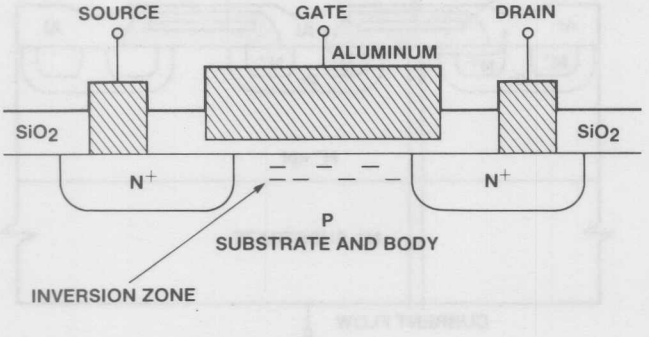

Figure 1. The Cross Section of a Conventional MOSFET

VMOS

Most of the deficiencies in the planar MOSFET's can be overcome by going to a structure that allows the current to flow vertically, and in which the channel length is controlled by diffusion processes rather than mask spacing. The VMOS structure shown in Figure 2a is a particularly good solution. A VMOS device fabrication process starts with an N^+ substrate with an N^- epi layer. A P region is diffused in and then an N^+ layer is diffused within the P region. Up to this point the process is very similar to that for a double diffused NPN transistor (Figure 2b), but, rather than applying the base and emitter metal, a V groove is anisotropically etched in the surface of the device, a silicon oxide insulating layer is grown, and finally, source and gate metal is deposited. Note that the source metal overlaps the P and N^+ regions so that the base and emitter of the NPN transistor are connected together.

By applying a positive potential between the gate and source, the P region close to the gate can be electrostatically inverted to N type material, and a conducting channel formed. Thus the source and gate connections are on the upper surface while the drain is on the bottom, and current flow is essentially vertical. In addition, the channel lengths are controlled by the diffusion processes and can be made quite short. This structure allows very efficient utilization of the silicon, and high power MOSFET's can be fabricated.

INTERSIL, INC., 10710 N. TANTAU AVE., CUPERTINO, CA 95014 (408) 996-5000 TWX: 910-338-0171

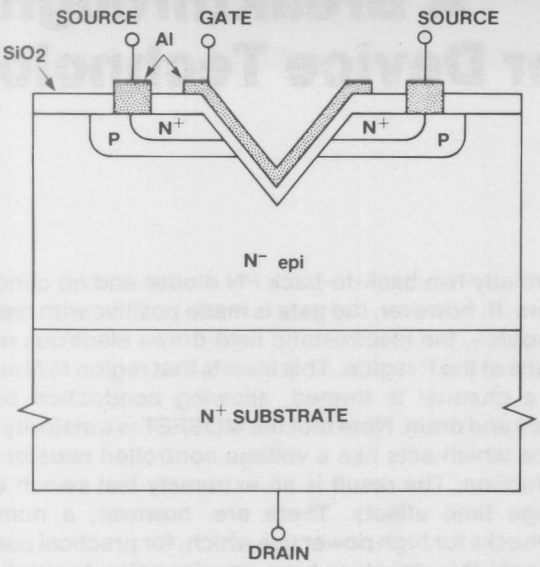

Figure 2A. Conventional Metal Gate VMOS Structure With Sharp "V" Groove

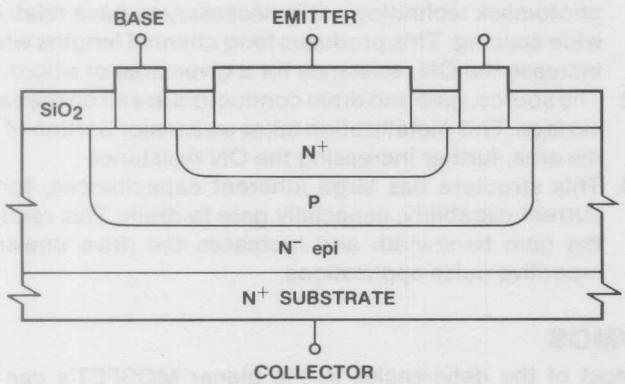

Figure 2B. Typical Double Diffused NPN Transistor

However, this basic structure has some drawbacks. The sharp bottom of the V groove produces a strong field concentration between the gate and drain, and there is also a tendency for the gate oxide layer to thin down around the tip of the V. The result is limited high voltage capability due to gate oxide breakdown even though the gate does not see the full drain-source voltage. The use of an aluminum gate can cause long term reliability problems due to ion migration (principally sodium) through the gate oxide which leads to variations in the device threshold voltage. A channel is formed on each side of the V groove, and if the groove does not penetrate well past the P region into the epi layer, it is possible to experience excessive current densities which may cause current injected avalanche breakdown in high power devices.

UMOS

Most of the VMOS problems can be relieved by using a flat bottomed groove with a combined silicon and aluminum gate structure as shown in Figure 3. The process is very similar to VMOS except that the etching is halted while the bottom of the groove is relatively wide. A layer of oxide is grown and overlaid with a layer of polycrystalline silicon

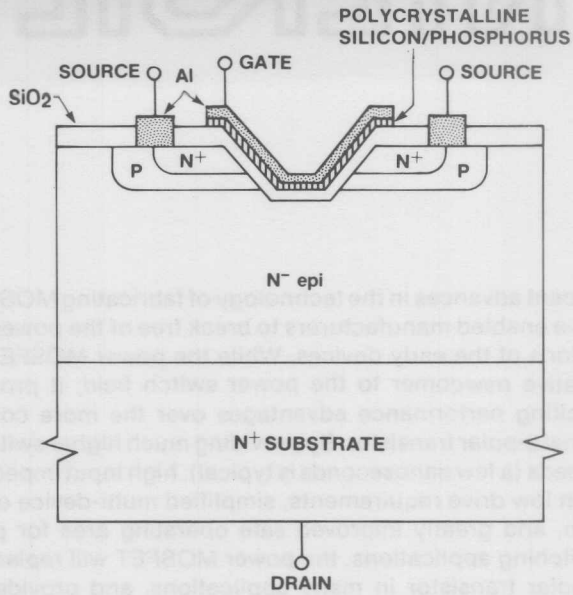

Figure 3. Flat-Bottomed Groove Silicon Gate VMOS

doped with phosphorus. Phosphorus doped poly-silicon is an effective ion migration barrier but it is not a particularly good conductor, having a resistance about 3000 times that of aluminum. In a large device, this could lead to a slow turn-on time due to the gate resistance. To alleviate this problem, a layer of aluminum is applied over the silicon-gate to provide high conductivity. Another benefit of the silicon-gate process is increased manufacturing yield which lowers the device cost. This is the standard Intersil process for VMOS devices.

Vertical DMOS

While the modified VMOS process is very effective for voltages under 150V, high field problems still exist, and the groove spacing requirements increase the die area. A vertical DMOS (double-diffused MOS) process has been developed to alleviate these problems. This structure is shown in Figure 4. The process begins as before (for an N-channel device) with an N^- epi layer grown on an N^+ substrate. Regions of P^- are then diffused, and inside these, regions of

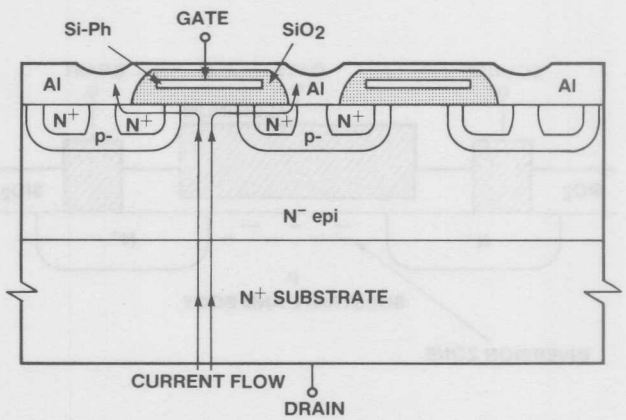

Figure 4. Vertical DMOS Structure

N^+. A silicon-gate is imbedded in SiO_2 and the source and gate metallization are then added to complete the device. The current flow is at first vertical and then horizontal, with the drain on the N^+ substrate.

This basic process has a number of different names, DMOS, TMOS, ZMOS, HEXFET®, etc. The processes are basically the same; the primary differences being in the geometry of the P and N regions and the interconnections. The HEXFET®, for example, uses hexagonal P regions, which allow a very low ON resistance by maximizing the channel perimeter. Unfortunately, as presently implemented, the silicon-gate structure has a very high series resistance which increases the switching time quite significantly. Intersil uses an alternate geometry that retains the low ON resistance but reduces the gate resistance.

Other Geometries

A variety of other geometries are in use to produce power MOSFET's and junction FET's.

Figure 5 shows a MOSFET structure developed by Hitachi.[1] The gate structure overlays the checkerboard of N and P regions to form the channels, and the N regions connect to the N^+ substrate so that the drain is on the back side of the die. To date, devices using this structure display a rather restricted Gain Bandwidth product of 0.6 to 1.5 MHz and a relatively high $r_{DS(on)}$ for a given die area. The primary application for these devices is audio amplifiers.

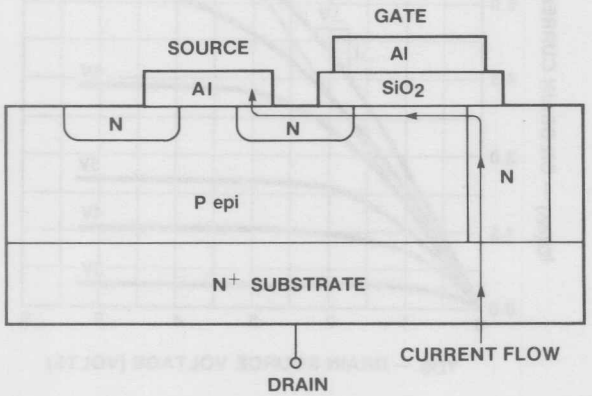

Figure 5A.

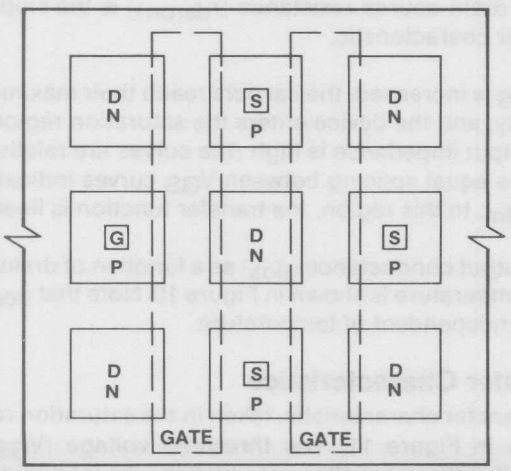

Figure 5B.

The SONY Corporation manufactures a device called a VFET,[2] which bears no relationship to the VMOS device, and is in fact, a vertical depletion mode junction FET. The structure, shown in Figure 6, produces a device with a square law transfer characteristic, while the output characteristic (Figure 7) is very much like a low μ triode. The disadvantages are relatively low stage gain, substantial gate current if the gate is driven positive, and the relatively high gate resistance and input capacitance which tend to reduce the gain bandwidth product.

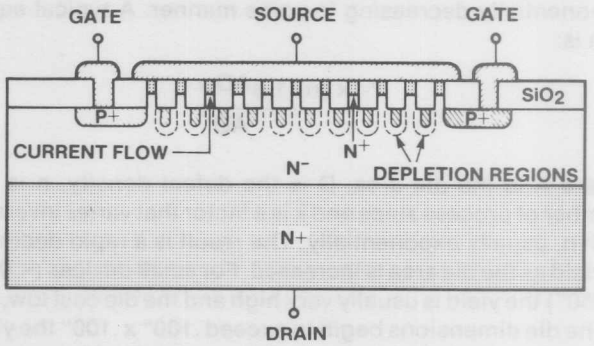

Figure 6. VFET Structure

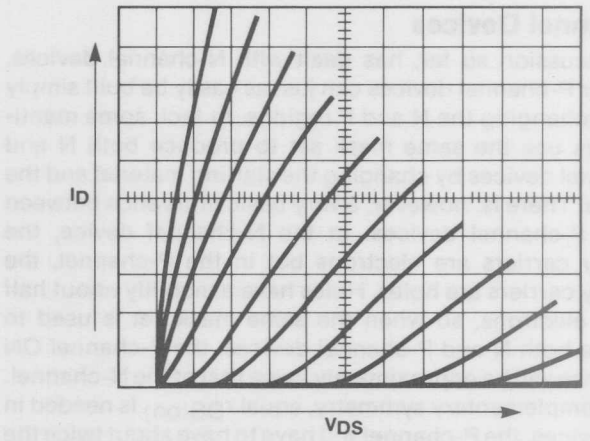

Figure 7. VFET Output Charactistics

BASIC LIMITATIONS

$r_{DS(on)}$ Versus Breakdown Voltage

As the breakdown voltage of a MOSFET (or a bipolar) is increased, the ON resistance, for a given die area and process, will increase exponentially by a factor of 2.3 to 2.7. [i.e., $r_{DS(on)} \propto k\ (BV_{DS})^{2.3-2.7}$] For example, if the breakdown voltage of a 1Ω, 100V device is increased to 200V, the die area must increase by five times to maintain the same ON resistance. The reason for this is twofold: first, the resistivity of the epi layer must be increased to raise the avalanche breakdown voltage. Second, the thickness of the epi layer must be increased to assure that the depletion region remains totally within the epi layer. Typically, in a 400V device, the epi resistivity will be 1500 to 2000 times greater than the N^+ substrate, so that at high voltages the ON resistance is dominated by the epi layer. The net result is that in either a MOSFET or a bipolar device, when the breakdown

voltage (for a given die area) is increased, the power handling capability is reduced. Conversely, if the same power capability is needed, then a larger die area is required.

Die Area Versus Yield/Cost

The ON resistance of a power FET is proportional to the die area. If a large die area is used to reduce the ON resistance, the number of dice per wafer will decrease; additional dice are lost due to inherent wafer defects and the increased scrap zone around the wafer center and periphery. The yield (Y) can be represented by a variety of equations, all of them exponentially decreasing in some manner. A typical equation is:

$$Y = k\,(n)\frac{(1-e^{AD})^2}{AD}$$

where A is the die area, D is the defect density, n is the number of process steps and k is a factor that varies inversely with n, usually exponentially. The result is a rapid decrease in yield as the die area is increased. For small devices (<.050" x .050") the yield is usually very high and the die cost low, but as the die dimensions begin to exceed .100" x .100" the yield drops and the cost per die increases rapidly. Many power FET's are larger than those dimensions, so for example, a one ohm 450 volt device may cost four to six times as much as a 2.5 ohm device with the same breakdown rating.

P-Channel Devices

The discussion so far, has dealt with N-channel devices, however P-channel devices can just as easily be built simply by interchanging the N and P regions. In fact, some manufacturers use the same mask set to produce both N and P-channel devices by changing the starting material and the process. There is, however, a very basic difference between N and P-channel devices. In the N-channel device, the majority carriers are electrons but in the P-channel, the majority carriers are holes. Holes have a mobility about half that of electrons, so when the same mask set is used to produce both N and P-channel devices, the P-channel ON resistance will be approximately twice that of the N-channel. If, for complementary symmetry, equal $r_{DS(on)}$ is needed in both devices, the P-channel will have to have about twice the area. The larger structure will also have more capacitance, so the devices will not be symmetrical in this respect, and in large devices the cost of the P-channel device will be higher.

LINEAR CHARACTERISTICS

Output Characteristics

The output characteristics for an IVN5200 are shown in Figures 8 and 9, where the drain current (I_D) is shown as a function of drain-source voltage (V_{DS}) with the gate-source voltage (V_{GS}) as a parameter. Two distinct regions of operation are apparent, the linear region and the saturated region. Be careful, these terms do not have the same meaning as they do for bipolar devices. In fact, they are almost exactly opposite! In the linear region ($V_{DS} \simeq$ 0-5V) the voltage across the channel is not sufficient for the carriers to reach their maximum drift velocity. In this region, the FET operates as a square law device; the static drain-source resistance [$r_{DS(on)}$] is equal to V_{DS}/I_D at each point and the small

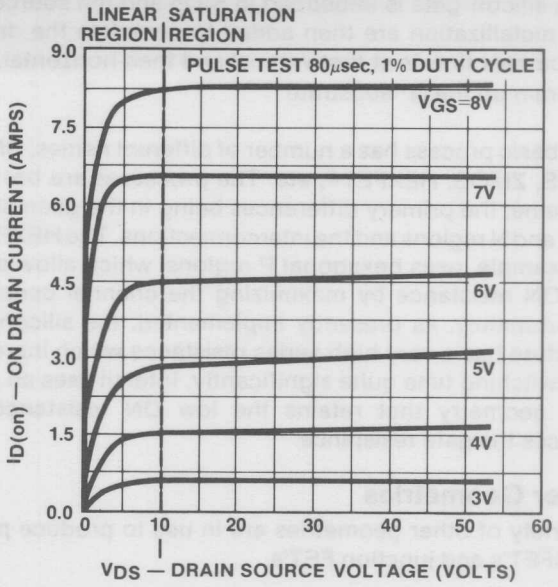

Figure 8. Output Characteristics IVN5200

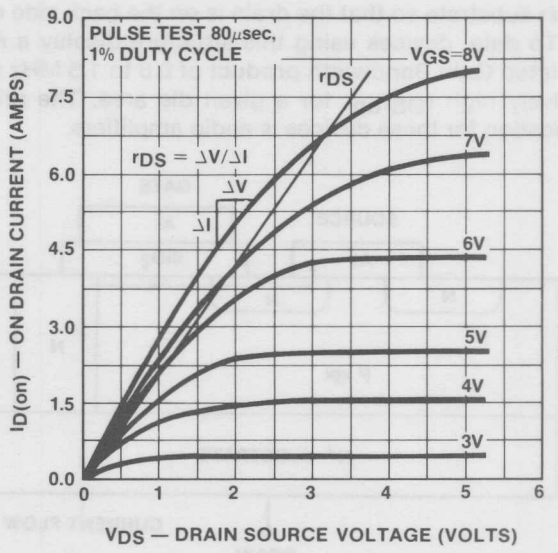

Figure 9. Output Characteristics IVN5200

signal drain-source resistance [$r_{ds(on)}$] is the slope of the transfer characteristic.

As V_{DS} is increased, the carriers reach their maximum drift velocity, and the device enters the saturation region where the output impedance is high (the curves are relatively flat) and the equal spacing between V_{GS} curves indicates constant g_m. In this region, the transfer function is linear.

The output conductance (g_{os}) as a function of drain current and temperature is shown in Figure 10. Note that g_{os} is relatively independent of temperature.

Transfer Characteristics

The transfer characteristic, taken in the saturation region, is shown in Figure 11. The threshold voltage [$V_{GS(th)}$] is defined as the gate voltage at which the device begins to turn on. The threshold voltage can be found by extending the

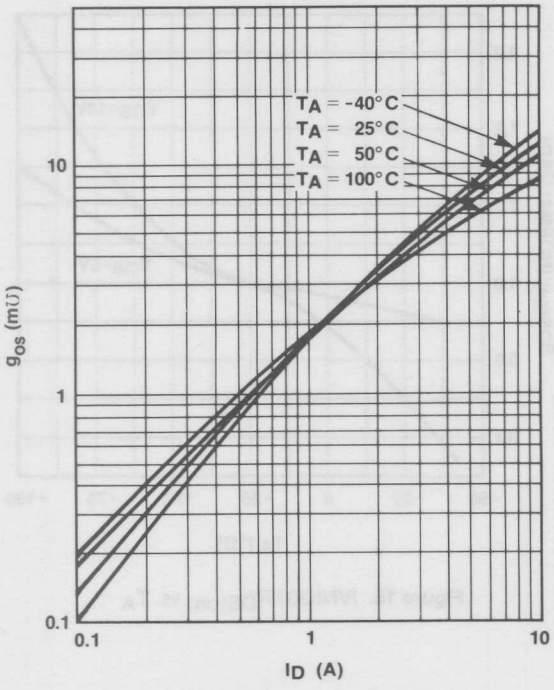

Figure 10. IVN5200 g_{os} vs. I_D

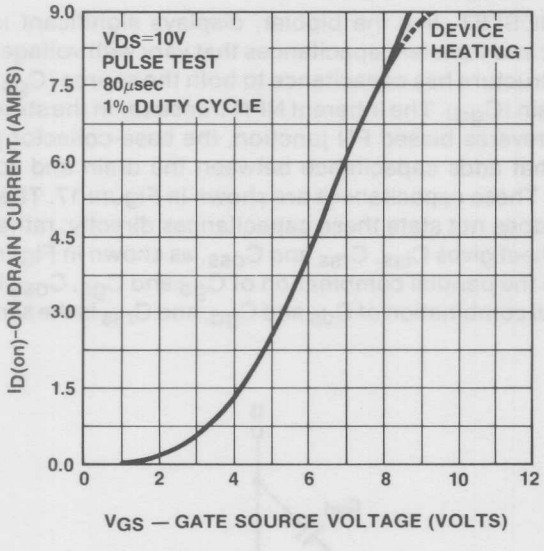

Figure 11. Transfer Characteristic

linear portion of the transconductance curve (Figure 14) to zero. As a practical matter however, because of the inconvenience of this measurement, manufacturers define $V_{GS(th)}$ as the gate voltage at which some small current, usually 1 to 10 mA, depending on the device size, begins to flow, and this value for $V_{GS(th)}$ is usually somewhat higher than the extrapolated intercept. The threshold voltage is a function of temperature; figure 12 shows this relationship for two different IVN5200's. The temperature coefficient is 6 mV/°C. Note that although the two devices have different threshold voltages, the temperature coefficients are nearly identical. This is typical within device types, and extremely helpful when paralleling multiple devices. For switching applications, the variation of $V_{GS(th)}$ with temperature is not very significant, but linear applications may require the bias point to be stabilized by using a source resistance or other negative feedback scheme.

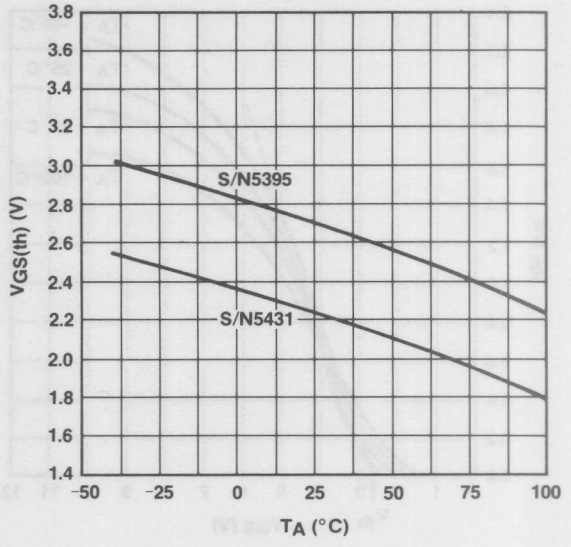

Figure 12. IVN5200 $V_{GS\ (th)}$ vs. T_A

The small signal forward transconductance (g_{fs}) curves are given in Figures 13 and 14. As shown in Figure 13, for a given V_{DS} the transconductance increases with I_{DS} until a point is reached where the transconductance is constant. This is characteristic of short channel MOS devices.

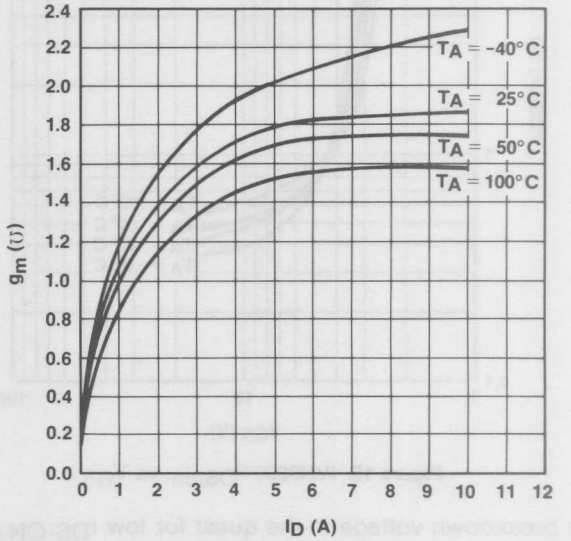

Figure 13. IVN5200 $g_m$vs. I_D

The $r_{DS(ON)}$ of a MOSFET is made up of two components, the channel ON resistance and the bulk resistance of the device. In low voltage devices (<100V) $r_{DS(ON)}$ is primarily limited by the channel resistance, but in higher voltage devices, the minimum value of $r_{DS(ON)}$ is dominated increasingly by the resistance of the epi layer. The channel resistance is controlled by the degree of gate enhancement, as shown in Figure 15. It can be seen that the ON resistance can be reduced by increasing V_{GS} to about 15 volts. However, a point of diminishing returns, sets in, so that little is gained by going above 15 volts. It is also important not to approach the

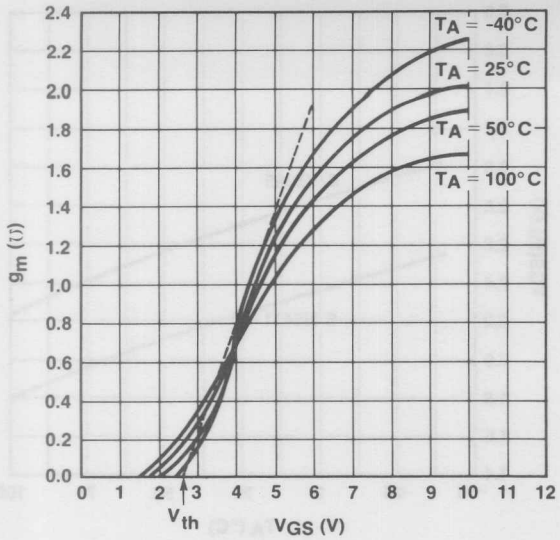

Figure 14. IVN5200 g_m vs. V_{GS}

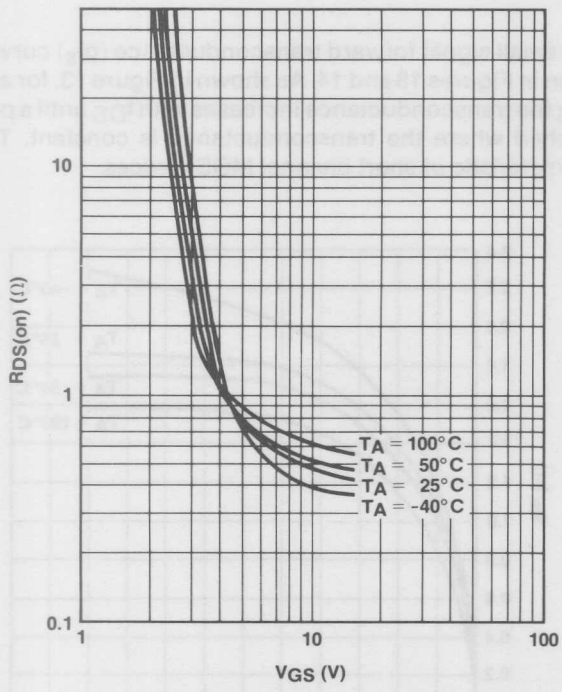

Figure 15. IVN5200 $R_{DS(on)}$ vs. V_{GS}

gate breakdown voltage in the quest for low $r_{DS(ON)}$. For higher $V_{GS(th)}$ devices, the curve will be displaced to the right but the shape remains essentially the same. Figure 16 shows the effect of temperature on $r_{DS(ON)}$ at different values of V_{GS}. Note that the temperature coefficient of $r_{DS(ON)}$ is positive. This positive temperature coefficient is a major reason for the improved safe operating area and ease of device paralleling. The value of this coefficient varies between +0.2 and +0.7%; this is caused by the competing effects of the positive TC of the silicon versus the negative TC of $V_{GS(th)}$. As shown by the curve for V_{GS} = 5V, when V_{GS} is close to $V_{GS(th)}$, the threshold effects predominate, and as the V_{GS} is increased the TC begins to take on the characteristics of the silicon. This example shows a low voltage (80V) device; in high voltage devices, r_{DS} is dominated more by the bulk resistance and the TC is more nearly linear, with typical values of +0.6 to +0.7%.

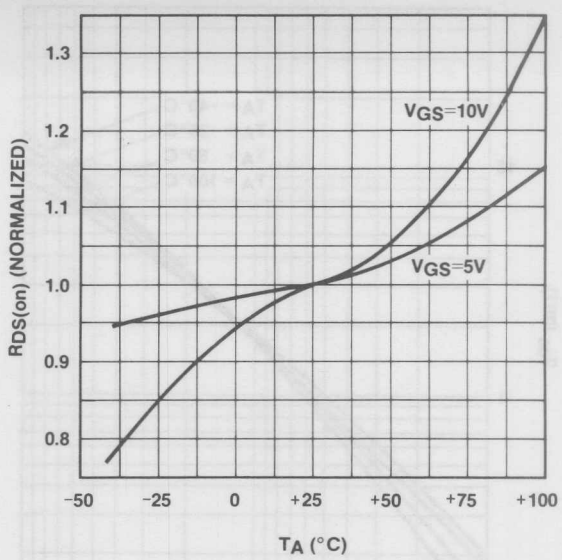

Figure 16. IVN5200 $R_{DS(on)}$ vs. T_A

Input Characteristics

The MOSFET, like the bipolar, displays significant input, output and transfer capacitances that vary with voltage. The gate structure has capacitance to both the source (C_{gs}) and the drain (C_{gd}). The inherent NPN transistor in the structure has a reverse biased PN junction, the base-collector junction, that adds capacitance between the drain and source (C_{ds}). These capacitances are shown in Figure 17. The data sheet does not state these capacitances directly; rather the data sheet gives C_{iss}, C_{rss} and C_{oss}, as shown in Figure 18. C_{iss} is the parallel combination of C_{gd} and C_{gs}, C_{oss} is the parallel combination of C_{ds} and C_{gd}, and C_{rss} is the same as C_{gd}.

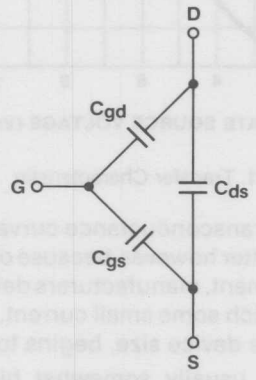

Figure 17.

As in any junction diode, the reverse biased capacitance decreases as the voltage is increased, so that C_{oss} has a much lower and nearly constant value above 10V than near zero. C_{iss} and C_{rss} are also reduced as V_{DS} is increased.

Because of the insulated gate structure, the input current (I_{GSS}) is normally very small, on the order of a few pico amperes at 25°C. This is made up of the leakage current through the gate structure, surface leakage current between

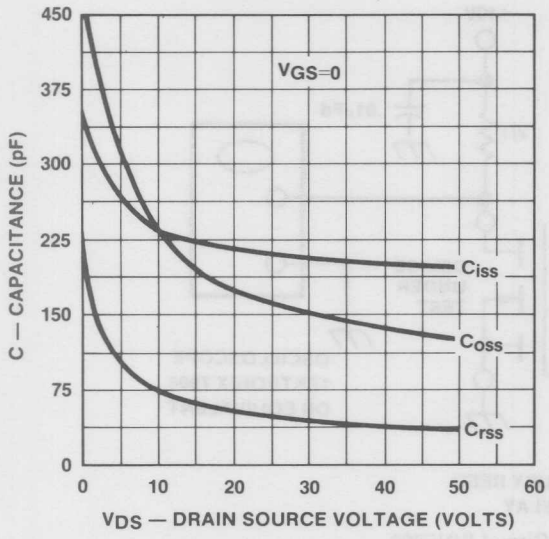

Figure 18. Capacitance vs. Drain-Source Voltage

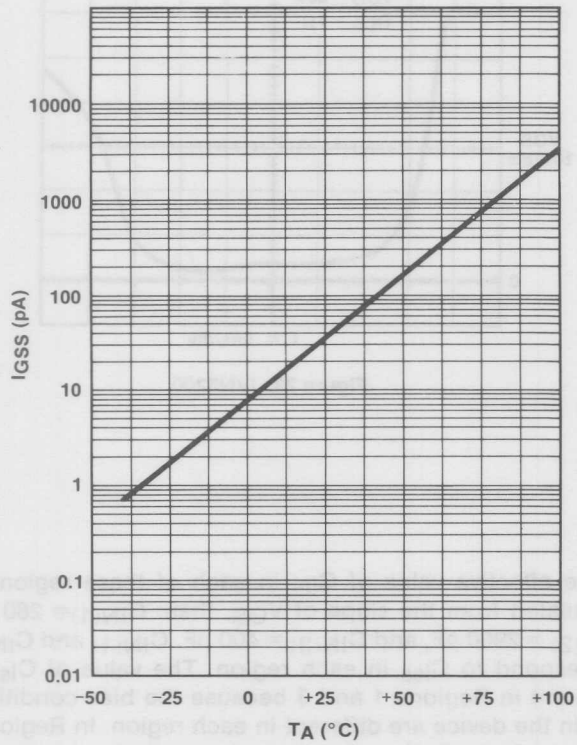

Figure 19. IVN5200 I_{GSS} vs. T_A

the package terminals, etc. Like any leakage current, I_{GSS} will increase exponentially with temperature; a typical characteristic curve is shown in Figure 19.

Occasionally, especially when an aluminum gate structure is used, the gate oxide can be contaminated with sodium ions. This produces a net positive gate charge, even when the gate and source are shorted, and results in a reduced threshold voltage. Unfortunately this is an uncontrolled process, and leads to random $V_{GS(th)}$ variations over life. Use of the silicon gate structure greatly reduces the possibility of sodium contamination.

Switching Characteristics

A major advantage of power MOSFET's is the very fast switching speeds of which they are capable. If one were able to charge the gate capacitance instantaneously, the switching time would be essentially the time it takes for the carriers to travel from the source to the drain. In present devices, this is about 50 to 200 picoseconds. Actual production devices can be switched in less than one nanosecond if a suitable pulse source is used to drive the gate. This can be done with a mercury wetted relay and a transmission line pulse source. For switching times longer than 5 to 10 ns, the turn-on time is limited primarily by the drive source resistance and the input capacitance (Figure 20a). As the switching times are reduced, the effect of parasitic inductances in the package and generator connections become important, (Figure 20b) so that it is difficult to turn on a TO-3 case device in much less than 2 - 3 ns.

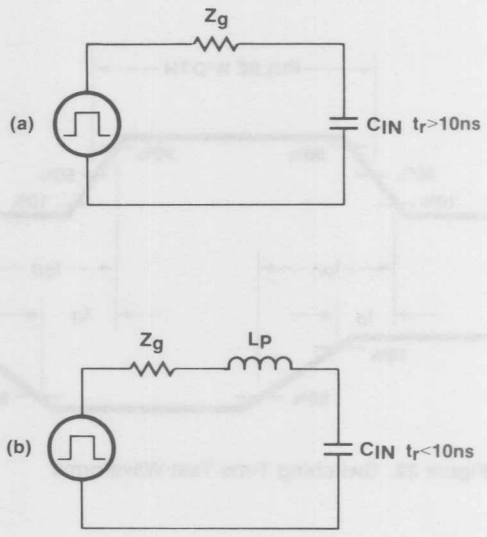

Figure 20.

The MOSFET switching times change very little with temperature, as opposed to bipolars. The idealized switching waveforms shown in Figure 22, presume the use of a specific test circuit like that shown in Figure 21. The actual transition times seen by the user may be greater or less than those shown on the data sheet, depending on the drive available. The delay at turn-on is due to the length of time it takes for the gate voltage to rise to $V_{GS(th)}$, where the device begins to conduct. In most switching applications, sufficient gate drive will be supplied to obtain the minimum $r_{DS(ON)}$; this corresponds to the area in Figure 15 where r_{DS} changes relatively slowly with V_{GS}. The result is a turn-off time delay where V_{GS} has to drop significantly before r_{DS} begins to rise.

The actual switching time test waveforms for a typical IVN5200 switching 60V at 8A are shown in Figure 23. The positive drain voltage spike at turn-on is due to the coupling of the drive pulse to the output by C_{rss} during the turn-on delay time. A similar negative pulse can occur during the turn-off delay time.

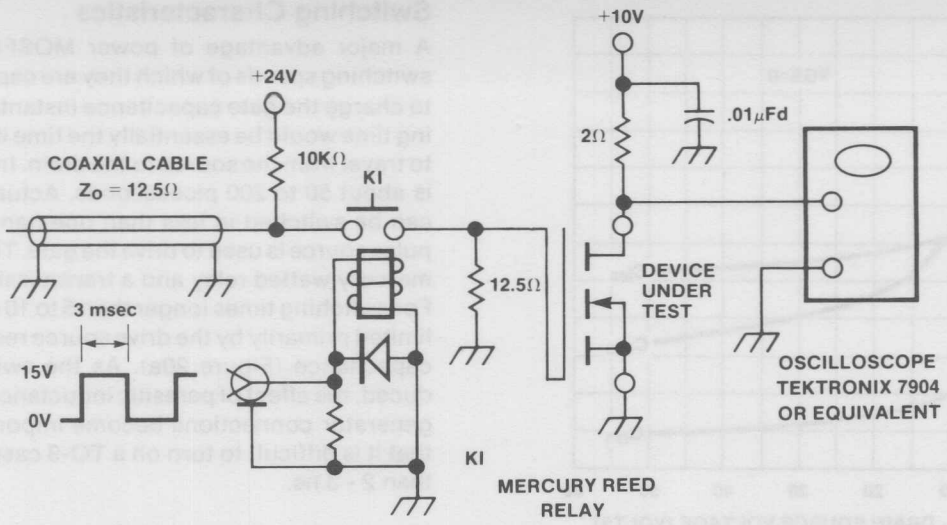

Figure 21. Switching Time Test Circuit IVN5200

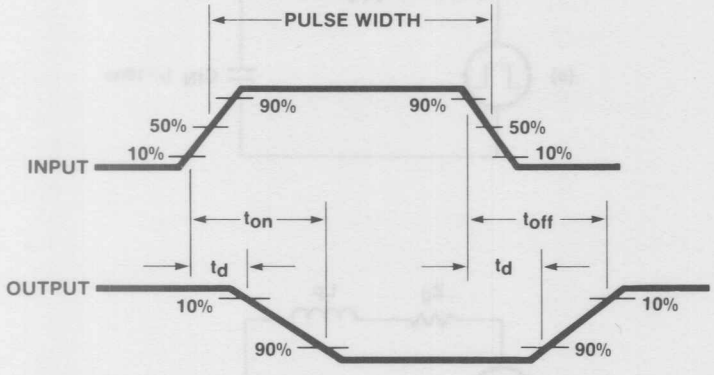

Figure 22. Switching Time Test Waveforms

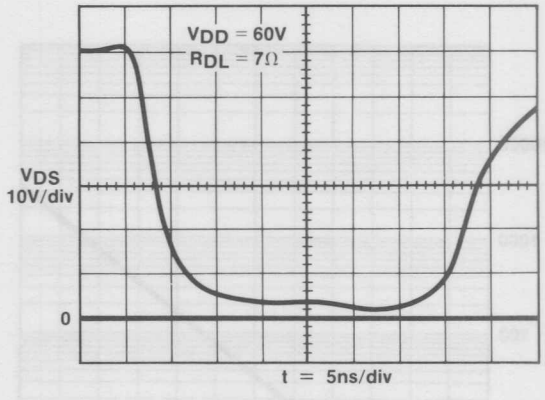

Figure 23. IVN5200

Input Capacitance

The actual gate input capacitance is highly non-linear, so that switching time and drive power calculations based on the average input capacity and the source resistance tend to be rather inaccurate. A better approach is to look at the gate charge as a function of V_{GS}. If the gate is driven from a current source (Figure 24), and the current integrated to give the charge, the curves in Figure 25 are obtained. For simplicity, the curves for V_{DD} = 60V are reproduced in Figure 26.

If one looks at V_{GS} as a function of charge, it is clear that three distinct regions exist and the dynamic input capacitance has a different value in each region. Region 1 corresponds to V_{GS} between zero and threshold, where the device is essentially off. The linearity of the voltage rise indicates that the capacitance is constant. In Region 2, the rate of rise of V_{GS} is sharply reduced, indicating a large increase in capacitance. This corresponds to the region where V_{DS} is falling and the Miller capacity appears at the input. In Region 3, the slope of V_{GS} is again increased, although not quite equal to that of Region 1, and is relatively linear. This corresponds to the region where the device is on and V_{DS} is no longer changing, so that the Miller effect is absent.

If the effective value of C_{IN} in each of these regions is calculated from the slope of V_{GS}, then: $C_{IN(1)}$= 260 pF, $C_{IN(2)}$ = 2900 pF, and $C_{IN(3)}$ = 400 pF. $C_{IN(1)}$ and $C_{IN(3)}$ correspond to C_{iss} in each region. The value of C_{iss} is different in Regions 1 and 3 because the bias conditions within the device are different in each region. In Region 2, $C_{IN(2)} \simeq C_{iss} - Av\, C_{rss}$, where $Av = \Delta V_{DS}/\Delta V_{GS}$, $Av\, C_{rss}$ is the Miller capacity.

The energy (W) required to turn the device ON is:

$$W = 1/2\ (\Delta V_G)\ (\Delta Q_G) \text{ watt-seconds}$$

If the gate is driven from a resistive source, ON and OFF repetitively, at a rate f_O then the drive power required is:

$$P = (\Delta Q_G)\ (\Delta V_{GS}) f_O$$

Driving an IVN5200 to a V_{GS} of 10V with f_O = 100 kHz, the drive power required is 7.5 mW. This is a vast improvement over a comparable bipolar device.

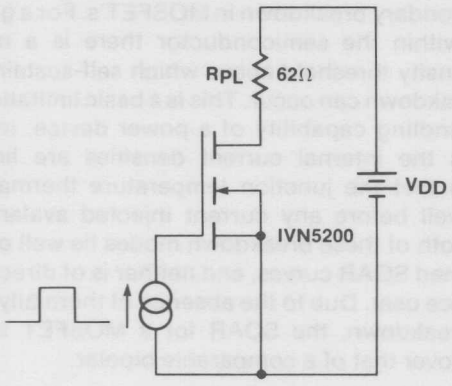

Figure 24.

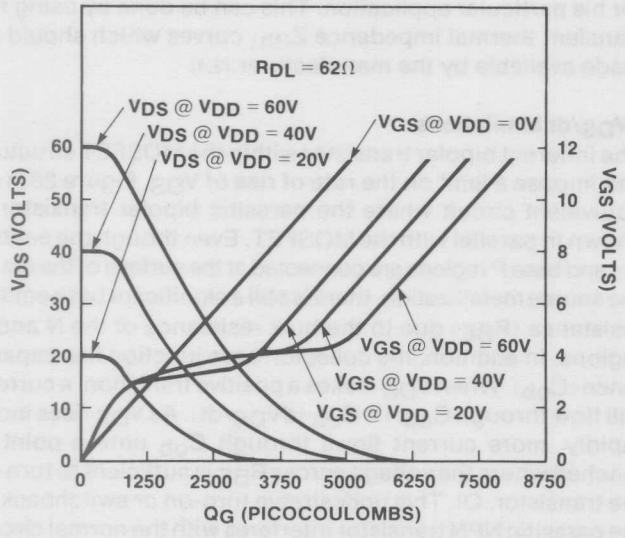

Figure 25. IVN5200 Dynamic Input Characteristics

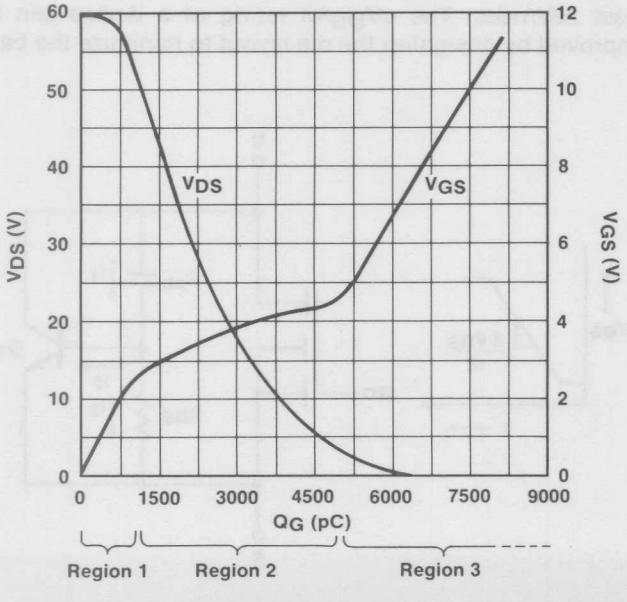

Figure 26. IVN5200 V_{DS} & V_{GS} vs. Q_G

Safe Operating Area

To achieve satisfactory service life from any power semiconductor, the circuit designer must assure that the device is operated within the voltage, current and thermal capabilities inherent in the particular device. To assist the designer, the manufacturer provides a table of maximum ratings, a safe operating area curve (SOAR) and a thermal impedance curve.

A typical absolute maximum ratings table is reproduced in Table I. While the information in the table is useful, it is not sufficient by itself for a power device. To adequately define safe operating conditions it is necessary to use the SOAR curve, like that reproduced in Figure 27. For a power MOSFET, the SOAR curve will have three boundary regions. Region 1 is defined by the breakdown voltage capability of

Table I. Absolute Maximum Ratings
(25°C unless otherwise noted)

Drain-source Voltage	
IVN5200TND, IVN5201TND	40V
IVN5200TNE, IVN5201TNE	60V
IVN5200TNF, IVN5201TNF	80V
Drain-gate Voltage	
IVN5200TND, IVN5201TND	40V
IVN5200TNE, IVN5201TNE	60V
IVN5200TNF, IVN5201TNF	80V
Continuous Drain Current (see note 1)	4.0A
Peak Drain Current (see note 2)	10A
Gate-source Forward Voltage	+30V
Gate-source Reverse Voltage	-30V
Thermal Resistance, Junction to Case	10°C/W
Continuous Device Dissipation at (or below)	
25°C Case Temperature	12.5W
Linear Derating Factor	100mW/°C
Operating Junction	
Temperature Range	-55 to +150°C
Storage Temperature Range	-55 to +150°C
Lead Temperature	
(1/16 in. from case for 10 sec)	+300°C

Note 1. T_C = 25°C; controlled by typical $R_{DS(ON)}$ and maximum power dissipation.

Note 2. Pulse width 80μsec, duty cycle 1.0%.

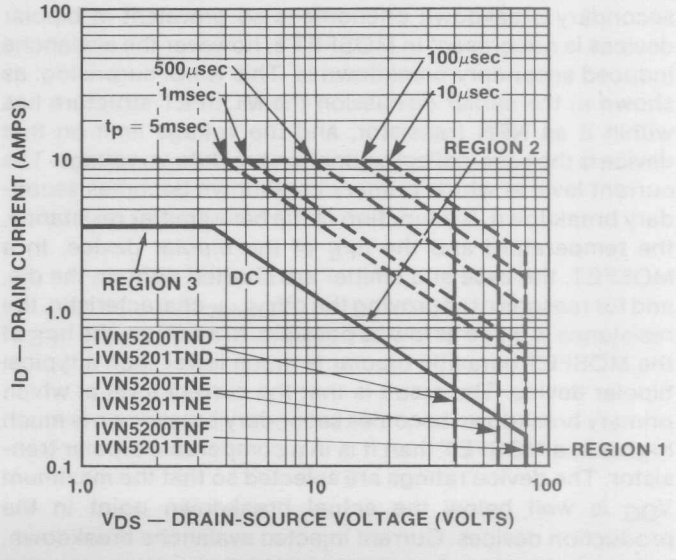

Figure 27. SOAR

the device. Region 2 is defined by the thermal capability of the device. Normally a maximum junction temperature of 150°C is specified, so that in Region 2 the power dissipation is limited to a *peak* junction temperature of 150°C. This results in Region 2 being defined by a family of curves that allow higher peak power for shorter pulse widths. Region 3 is defined by the current capability of the device. The current capability of a given device may be limited by the bond wire diameter, the area of the bonding pad on the die, or by the metallization on the die surface. Whereas the breakdown voltage and junction temperature limitations can be readily determined by direct measurement, the current limitations are empirically derived from life testing. The maximum current is limited to a value which has been found to give an acceptable service life. In a bipolar transistor, the rapidly decreasing h_{FE} at high currents effectively discourages operation in excess of the current ratings. In a MOSFET, however, the gain is not reduced at high currents, and there may be a temptation, in fast pulse applications where a high drain-source voltage may be acceptable, to operate with very short high current pulses in excess of the ratings. Even if the device dissipation is very low, this is inadvisable for two reasons. First, the reliability or service life of the device is undefined and likely to be shortened, and second, if the current density in the device is increased sufficiently, it is possible to reach the level where current injected avalanche breakdown occurs, possibly destroying the device.

For a bipolar transistor the SOAR curves will have a fourth boundary. This region is defined by the thermally induced second breakdown characteristic. In a bipolar device, there are several ways to induce secondary breakdown. The first is thermal, where the negative temperature coefficient of the conduction resistance causes localized hot spots to be formed. When the temperature of a hot spot is sufficiently high, it's impedance is drastically reduced, funneling the collector current through a small area and usually destroying the device. Another mechanism for inducing secondary breakdown is via avalanche breakdown. If the collector voltage is raised to the breakdown point of the collector-base junction and a significant current allowed to flow, the device will go into secondary breakdown. It has been widely advertised that MOSFET's do not exhibit secondary breakdown. This is not true. It is generally accepted that at normal junction temperatures, the thermally induced secondary breakdown phenomena so prevalent in bipolar devices is not present in MOSFET's, however the avalanche induced secondary breakdown is. This is not surprising; as shown in the earlier discussion the MOSFET structure has within it an NPN transistor, and the voltage limit on that device is the base-collector junction breakdown voltage. The current level at which primary breakdown becomes secondary breakdown, is a function of the base emitter resistance, the temperature and the h_{FE} of the bipolar device. In a MOSFET, the base and emitter are shorted right on the die, and for reasons of improving the dV_{DS}/dt characteristic, the resistance is made as low as possible. In addition, the h_{FE} of the MOSFET parasitic bipolar is much lower than a typical bipolar device. The result is that the current level at which primary breakdown becomes secondary breakdown is much higher in a MOSFET than it is in a comparable bipolar transistor. The device ratings are selected so that the maximum V_{DS} is well below the actual breakdown point in the production devices. Current injected avalanche breakdown, present in bipolars during reversed bias operation, can also lead to secondary breakdown in MOSFET's. For a given field gradient within the semiconductor there is a maximum current density threshold above which self-sustaining avalanche breakdown can occur. This is a basic limitation on the current handling capability of a power device. In present MOSFET's the internal current densities are limited by design, so that the junction temperature thermal limit is reached well before any current injected avalanching is present. Both of these breakdown modes lie well outside of the published SOAR curves, and neither is of direct interest to the device user. Due to the absence of thermally induced second breakdown, the SOAR for a MOSFET is greatly expanded over that of a comparable bipolar.

The normal manufacturers' SOAR curve is for a case temperature of 25°C and either DC or a *single* pulse. In the real world, of case temperatures above 25°C and repetitive pulses, the designer must modify the standard SOAR curves for his particular application. This can be done by using the transient thermal impedance $Z_{(th)}$ curves which should be made available by the manufacturer.[3,4]

dV_{DS}/dt Limitations

The inherent bipolar transistor within the MOSFET structure can impose a limit on the rate of rise of V_{DS}. Figure 28 is an equivalent circuit where the parasitic bipolar transistor is shown in parallel with the MOSFET. Even though the emitter N+ and base P regions are connected at the surface of the die by the source metallization, there is still a significant base emitter resistance (R_{BE}) due to the bulk resistance of the N and P regions. In addition, the collector-base junction has capacitance (C_{ob}). When V_{DS} makes a positive transition, a current will flow through C_{ob}; $I = C_{ob}(dV_{DS}/dt)$. As V_{DS} rises more rapidly, more current flows through C_{ob} until a point is reached where the voltage across R_{BE} is sufficient to turn on the transistor, QI. This undesirable turn-on or switchback of the parasitic NPN transistor interferes with the normal circuit operation and can destroy the device. The threshold for switchback varies widely from one device type and manufacturer to another. Maximum dV_{DS}/dt information is not yet a standard entry on the data sheets so the user will have to either consult with the manufacturer or test the devices himself. The IVN5000 and 5200 series parts will accept at least 20V/nsec. The dV_{DS}/dt rating of a device can be improved by designing the die layout to minimize the base

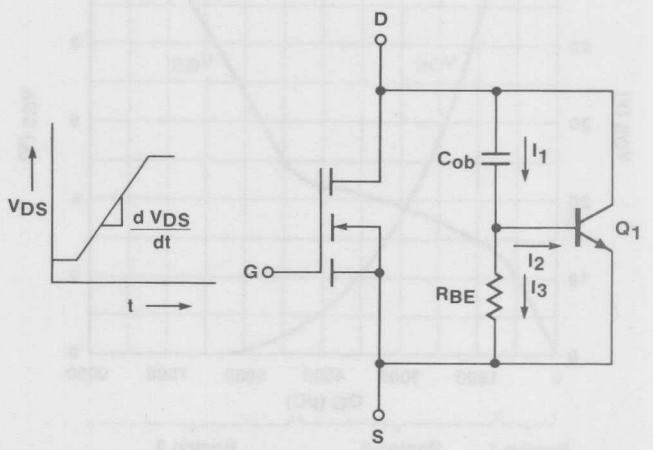

Figure 28. Parasitic Bipolar Switchback Equivalent Circuit

emitter resistance, and by controlling the P region doping to produce a bipolar with low h_{FE}. As in any bipolar transistor, the h_{FE} of the parasitic NPN is a function of temperature and the dV_{DS}/dt snapback threshold will decrease at high temperatures.

Use of the Internal Diode

For values of V_{DS} risetime below the threshold of switchback, the internal NPN transistor is inactive, and acts simply as a bipolar with the base shorted to the emitter. The equivalent circuit could be redrawn, as shown in Figure 29, with a diode in parallel with an ideal MOSFET. If V_{DS} is reversed, this diode will conduct and may be used in switching circuits as a rectifier or inductive energy clamp. The forward current and breakdown voltage ratings of this diode are equal to the current and voltage ratings of the parent MOSFET. The reverse recovery time t_{rr} of this diode can be excellent; the IVN500 and 5200 series devices have a typical t_{rr} of 60 to 70 ns. There are several reasons for the good recovery time of this diode. First, the MOSFET fabrication process produces essentially lightly doped epitaxial diodes with sharply defined doping gradients. Second, the P region is often formed by ion implanting, a process that causes dislocations in the crystal structure. These dislocations can act as recombination centers for the junction stored charge, thereby reducing the recovery time. The diode recovery time is senstitive to the process used and can vary widely from one manufacturer to another. Since t_{rr} is not, at present, a data sheet entry, the user is advised to contact the manufacturer for values.

The FET itself may be used as a synchronous rectifier. Once the channel has been formed by making V_{GS} positive, current will flow through the device in either direction with equal facility. For synchronous rectifier operation, V_{GS} is zero during the period when the drain is positive with respect to the source; when V_{DS} reverses, the gate is enabled and current flows through the device in the reverse direction. The forward drop is proportional to $r_{DS(ON)}$ and the current flowing; when the threshold of the parallel diode is reached the current will bypass the FET channel and the rectifier will act like a normal PN junction diode.

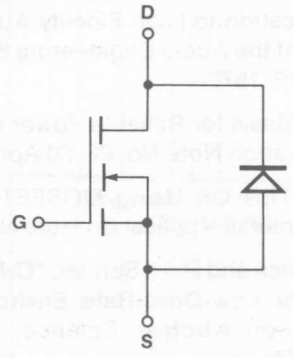

Figure 29.

For currents below the 0.6V diode threshold, the FET acts as an ultra high speed, high voltage, low capacity rectifier with no minimum offset voltage. This type of operation has been used for a power rectifier at 15 MHz. Presently available MOSFET's can be used in this manner, with the only drawback being the relatively high $r_{DS(ON)}$ values which restrict the current. In the future, it is reasonable to expect that low voltage, low $r_{DS(ON)}$ devices will be designed specifically for this service, which will make possible the efficient generation of large amounts of 2V or 5V power.

Internal Zener Protection

The first VMOS devices to appear on the market included an on-chip zener diode from gate to source to prevent gate breakdown due to static charging. This was formed by diffusing in an additional NPN transistor (Figure 30), with the zener action accomplished by the reverse breakdown of the base-emitter junction. The zener formed in this manner has a number of drawbacks. First, the power dissipation is very limited, 2 mA DC being typical, and second, if the gate is pulled negative to -0.6V or more, the NPN transistor will turn ON and draw current from the drain circuit. This may destroy the transistor. The result is poor reliability in many applications. Experience has shown that the gate structure of a power MOSFET is much more rugged than the gate structure in MOS IC devices, and given reasonable care in packaging, handling and installation there is no need for this zener in most applications.

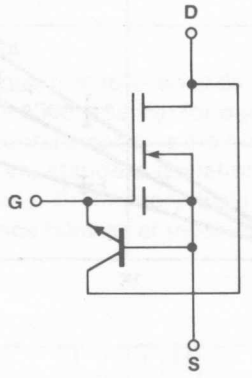

Figure 30. Parasitic NPN Transistor with Internal Zener

Multiple Device Operation

The positive temperature coefficient of $r_{DS(ON)}$ is of great assistance when MOSFET's are paralleled. It is possible, in DC applications, to parallel devices without any matching, and those devices that initially draw the most current will heat up and shift the current to other devices to more equally distribute the current. This is exactly the opposite of the scenario for bipolar devices. While the paralleling of unmatched devices will work, it is a poor idea because in DC applications a higher than necessary dissipation may occur and in switching applications it is possible for one device to turn on or off, before or after the other devices and have to accept the full load current. This may force the device to function outside of its safe operating area.

It is recommended that the user parallel devices which have been matched for $V_{GS(th)}$ to within 5 to 10% accuracy. This will assure that the turn-on and turn-off delays due to the gate voltage rise and fall times, relative to $V_{GS(th)}$, are nearly equal. $r_{DS(ON)}$ will also be matched, so that excessive differential heating is not required. Some additional improvement in $r_{DS(ON)}$ matching may be achieved by providing a higher V_{GS} drive so that all of the parallel devices run at their minimum $r_{DS(ON)}$ value.

It should be kept in mind that many MOSFET devices have gain bandwidths over 500 MHz. It is entirely possible for these devices to oscillate at very high frequencies, especially if multiple parallel devices are used. This is often an unsuspected cause of device failure. If the circuit designer is using a low bandwidth oscilliscope during breadboard development, it is possible not to be aware of the self oscillation therefore, use of an oscilliscope with a bandwidth of at least 200 MHz is recommended. The tendency towards oscillation can be greatly reduced by inserting ferrite beads or low value (50 - 100 ohm) resistors in series with the gate leads as shown in Figure 31.

For fast pulse applications, it is not sufficient to just match the devices. The circuit must also be reasonably symmetrical so that identical drive voltages are applied to each gate. At high speeds the inductive, as well as resistive, effects must be considered. If, for example, the parasitic inductance in the drain lead of Q_1 is much smaller than that in the drain lead of Q_N, when the devices are first turned ON most of the load current will initially flow through Q_1, even if the gate drives and threshold voltages are identical.

Variations in the stray gate circuit capacity and the device input capacity can also cause uneven turn-on in fast pulse applications. The effect of this can be reduced by equalizing the stray capacity and minimizing the intergate impedances. For applications requiring switching times below 5 to 10 ns., it may be necessary to match the device capacitances.

It should be kept in mind that most of the complications to paralleling mentioned above generally apply only for very fast pulses. Most applications will use transition times well above 10 nsec, where simple threshold voltage matching is all that is required. This is quite different from bipolar transistors, where paralleling more than two devices can become quite complex and expensive. Due to the practical difficulties of paralleling large numbers of individual bipolars, (and SCR's also), the trend has been to develop ever larger single devices. As the die size increases, a point is reached where either the thermal capabilities or the available mounting areas in the low cost packages are exceeded so that an expensive package is required. The heat sink will now see a concentrated thermal input through the relatively small package to heat sink contact area. The efficiency of a heat sink is better if the heat input is distributed in several sources rather than in one. While the arguments for using a single large bipolar device instead of multiple smaller devices are well founded, it does not necessarily follow that the same technique should be imitated in power MOSFET devices, especially in the light of the rapid cost increase of the larger die. The tradeoff may well be multiple small devices with low die and package and moderate heat sink costs, versus single devices with high die, package and heat sink costs. Just where the cost crossover point between single and multiple devices will be is yet to be determined, especially since the very high prices for the present larger MOSFET chips is certain to be reduced substantially. In any case, the user should be aware of this tradeoff which is quite different from bipolars, and not unnecessarily perpetuate the bias against parallel bipolars into the application of power MOSFET's.

Similar considerations apply if devices are to be operated in series. It is particularly important that all devices in the same series string come ON simultaneously, otherwise one device may take all of the voltage momentarily. The devices should be well matched for $V_{GS(th)}$ and careful attention given to producing simultaneous gate drive. In those applications where the maximum possible voltage capability is desired, it may be necessary to match the $r_{DS(ON)}/V_{GS}$ characteristic to assure equal voltage distribution during switching transitions. This is a much more complex procedure than $V_{GS(th)}$ matching, and should be considered only as a last resort. Analogous to the parasitic inductance in parallel operation is the effect of circuit parasitic capacitance in series operation. If differential drain-source capacitances (either in the device or in the circuit layout) exist in the series string, the transient voltages may not be shared equally.

RADIATION EFFECTS

The effects of nuclear radiation on power MOSFET's is beginning to receive a good deal of attention, although very little information has been published to date. Mr. John Buck, a radiation effects specialist with Litton Guidance and Control Systems, has provided the following information:

". . . When used in military systems with a nuclear hardening specification, VMOS and bipolar power devices have differ-

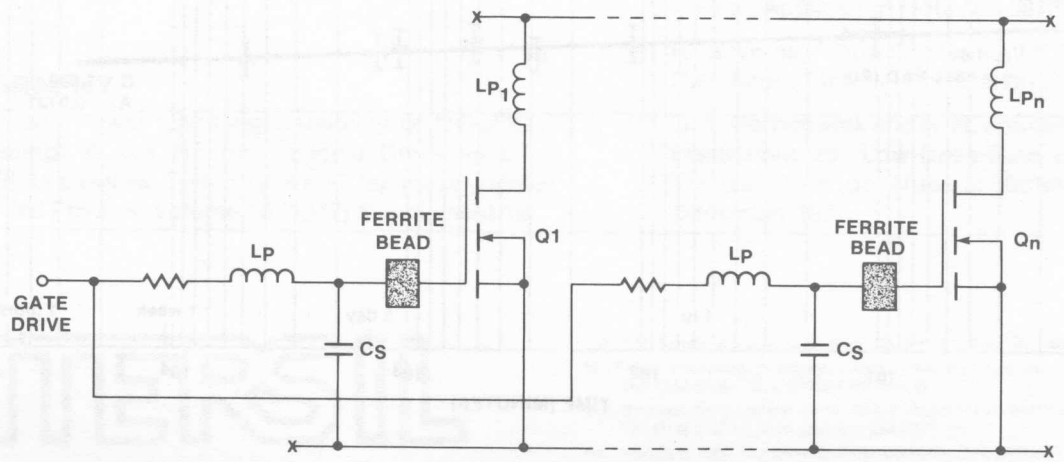

Figure 31. Parallel Devices

ent susceptibilities to the neutron and total dose radiation environments. VMOS devices are almost immune to neutron damage because they are majority carrier devices. Bipolar power devices, on the other hand, suffer considerable gain degradation in even moderate neutron environments. VMOS devices, however, are more adversely affected by the total dose (gamma) environment than are bipolar devices.

"Total dose damage to MOS devices occurs as a result of (the) trapped charge in the gate oxide layer. Charge is created when gamma particles collide with (atoms) in the oxide layer causing formation of electron — hole pairs. The electrons are more mobile than the holes, and are (more) easily swept from the oxide layer. The holes, on the other hand, tend to drift toward trapping sites where positively charged hole layers are formed. This positive charge remains in the oxide layer and causes the gate threshold voltage of the device to decrease. Total dose damage to MOS devices is measured as a shift in gate threshold voltage. Drain leakage current also increases with total dose as a result of the gate charge buildup.

"Two other factors are important in discussing total dose damage effects in MOS devices. One is the amount of gate bias applied during irradiation and the other is the effect of annealing. The more positive (the) gate bias applied during irradiation the greater the total dose damage. This is because the holes will migrate to trapping sites closer to the sensitive Si-SiO_2 interface.

"Annealing, on the other hand, is a curing effect since the process of annealing is the recombination of holes with electrons with (the) resulting decrease of trapped positive charges. Annealing will take place at room temperature, and the rate of annealing will increase at higher temperatures due to increased hole de-trapping.

"Derbenwich[5] has studied the combined effects of total dose damage and annealing. His method involves measuring gate threshold voltage shift versus total dose to determine a linear damage region for the device. Then another set of the same device type is irradiated to a dose level within the linear damage region and then allowed to anneal at room temperature under fixed bias for several weeks or longer. A transient annealing function is then determined for the device. His theory then states that the convolution of dose rate versus time with the transient annealing function predicts the net shift in gate threshold voltage when annealing is occuring simultaneously. This model provides a means for calculating MOS damage in low dose rate long term environments where annealing is a significant factor. A typical long term ionizing radiation environment is a satellite space application where the system may be irradiated and anneal simultaneously for a five year mission."

Test Results

"Figure 32 shows that total dose damage to IVN5000 parts is linear to about 5000 rads (Si) for bias voltages of 0, 2.5V and 5V. The horizontal error bars are dosimetry error estimates. Vertical bars are standard deviations based on a sample of five parts tested. Preliminary tests at higher bias voltages indicated device failures at much lower dose levels. Device

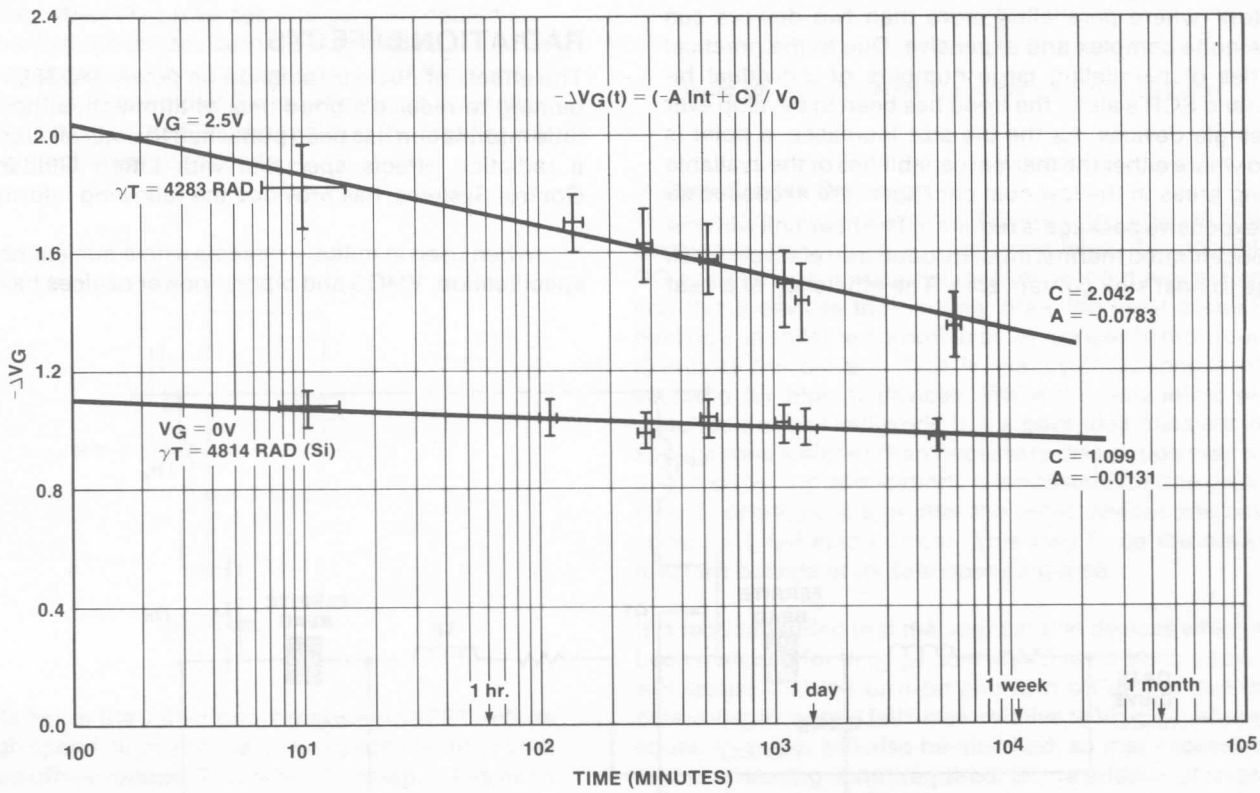

Figure 32. VMOS LINAC Annealing Curves

failure on these tests was due to leakage currents approaching or exceeding the drain currents at which the DC gate voltage measurements were made (500μA). Pulse testing at higher currents may have allowed gate voltage measurements to be made at slightly higher dose levels.

"Figure 33 shows the results of the annealing measurements made for 0V and 2.5V biased parts after LINAC irradiations to 4814 rad (Si) and 4283 rad (Si) respectively. It can be seen from Figure 33 that these total dose irradiations are within the linear damage regions for the device."

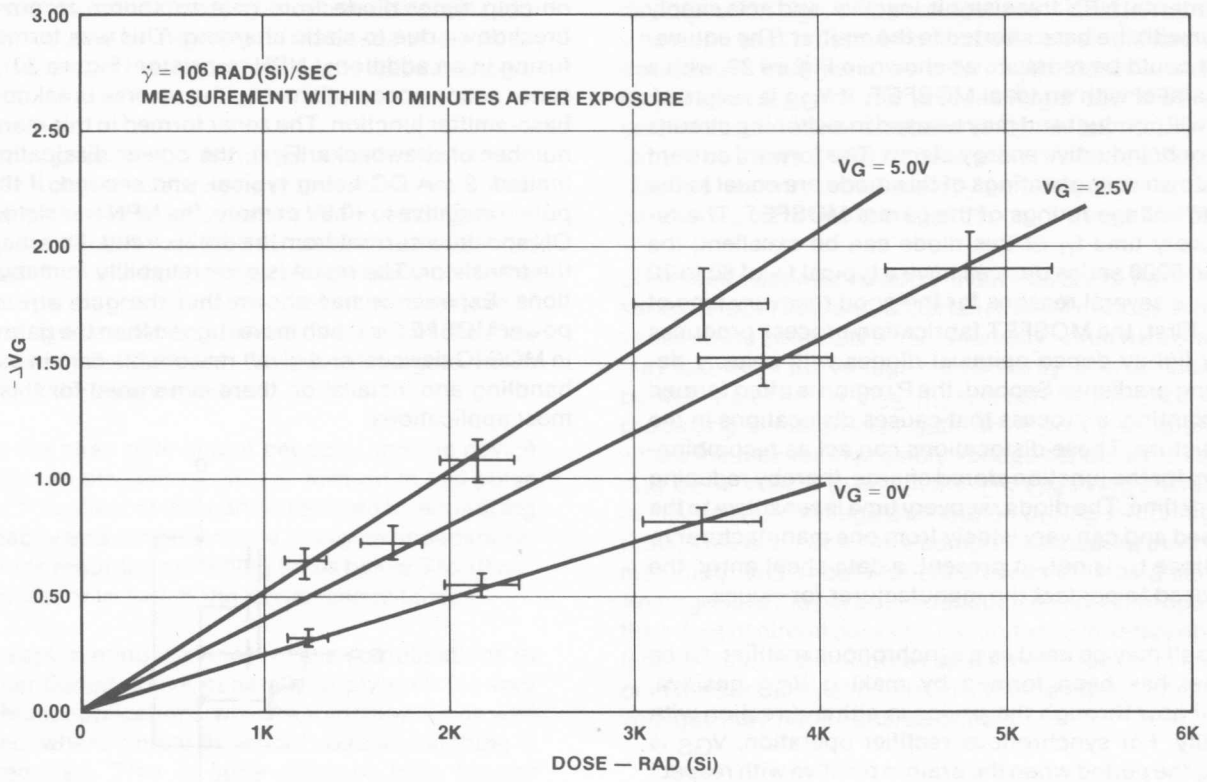

Figure 33. VMOS LINAC Test Results

ACKNOWLEDGEMENT

The author would like to thank B. Rosenthal, N. Zommer, and B. Wier for their assistance in preparing the discussion on structures and B. Smithson who did the characteristics curves.

BIBLIOGRAPHY

[1] M. Nagata, "Power Handling Capability of MOSFET," Proceedings of the 8th International Conference on Solid State Devices, Tokyo 1976 and Japanese Journal of Applied Physics, Volume 16 (1977) Supplement 16.

[2] H. Suwa and A. Ishitani, "Vertical Field Effect Transistor and It's Application to High Fidelity Audio Amplifiers," Proceedings of the Audio Engineering Society Convention, May 13-16, 1975.

[3] "SOAR—The Basis for Reliable Power Circuit Design," Phillips Application Note No. 68, 30 April 1975.

[4] R. Severns, "Tips On Using MOSFET's in Switching Regulators," Intersil Application Note No. A0-36.

[5] G. F. Derbenwick and H. H. Sander, "C-MOS Hardening Predictions for Low-Dose-Rate Environments," IEEE Transactions on Nuclear Science, *NS-24,* No. 6, December 1977.

INTERSIL

10710 N. Tantau Avenue, Cupertino, CA 95014 U.S.A., Tel: (408) 996-5000, TWX: 910-338-0171
9th Floor, Snamprogetti House, Basing View, Basingstoke, RG21 2YS, Hampshire, England, Tel: 0256-57361, TLX: 847227 INTRSL G
217, Bureaux de la Colline, de St. Cloud, Batiment D, 92213 Saint-Cloud, Cedex, France, Tel: (1) 602-57-11, TLX: Datelem 204280F
Bavariaring 8, 8000 Munchen 2, West Germany, Tel: 89/539271, TLX: 5215736 INSL D